

# Digitization of Documentary Heritage Collections in Indic Language Comparative

*Study of Five Major Digital Library Initiatives in India*

**Anup Kumar Das**

*Centre for Studies in Science Policy, School of Social Sciences, Jawaharlal Nehru University, New Delhi, India, anupdas2072@gmail.com, anup\_csp@mail.jnu.ac.in*

## Abstract

*Documentary heritage collections in Indic languages have been soul of indigenous digital libraries in South Asia. Some of the digital preservation initiatives in India received global acceptance are namely, Digital Library of India, Panjab Digital Library, Kalasampada Digital Library – Resource for Indian Cultural Heritage, National Databank on Indian Art and Culture, Traditional Knowledge Digital Library, and National Mission for Manuscripts, due to uniqueness in their collections and approaches. These projects also help in preserving socio-linguistically diverse cultural contents and achieving a sense of unity while online accessing using common platforms. This paper evaluates enrichment of collections and effectiveness of online platforms of five major digital library initiatives in India.*

## Author

Dr. Anup Kumar Das is an advance researcher and library professional attached with the Centre for Studies in Science Policy at Jawaharlal Nehru University (JNU), India. He was a UNESCO Consultant in the areas of Information for All and Open Access to Knowledge at its New Delhi office. He was awarded PhD degree from the Jadavpur University, India in 2008 for his doctoral thesis titled “*An Evaluative Study of Some Selected Libraries in India undergoing the Process of Digitization*”. He has published a number of papers in different journals and presented papers in different international/ national conferences. His recent work includes ‘*Open Access to Knowledge and Information: Scholarly Literature and Digital Library Initiatives – the South Asian Scenario*’, UNESCO, New Delhi, 2008. A list of his publications and presentations is available at <http://anupkumardas.blogspot.in>.

## 1. Introduction

India is the country of ‘unity in diversity’, where multicultural society embraces a diverse people with anthropologically different linguistic, religious, ethnic, demographic and regional backgrounds. Here cultural diversity and cultural pluralism are coexisting for the centuries. India is the country of origin of many legendary ancient literature that form rich collections of Asian documentary heritage. South Asia also plays a significant role in shaping up world literary traditions.

Over the time Indian cultural institutions became the repositories of rich collections of cultural heritage resources embracing culturally and linguistically diverse communities across states of India.

While traditional knowledge of linguistically diverse communities is largely un-documented, there were several attempts to collate them. Systematic documentation of traditional knowledge is centuries old practice of scholars and researchers to make knowledge re-usable by future the generations. These documentation initiatives ended up with producing literature of various kinds. On the other hand, some of the documentary heritage resources available with Indian institutions are on the verge of extinction due to lack of preservation and conservation initiative at the institutional level.

As a member country of UNESCO, India became de-facto signatory of the *UNESCO Universal Declaration on Cultural Diversity*, adopted unanimously by the UNESCO General Conference at its 31st session held on 2 November 2001. This is an international standard-setting legal instrument, which raises cultural diversity to the rank of “common heritage of humanity” [UNESCO, 2001]. The Declaration attempts to respond to two major concerns: (i) to ensure respect for cultural identities with the participation of all peoples in a democratic framework, and (ii) to contribute to the emergence of a favourable climate for the creativity of all, thereby making culture a factor of development.

The Article 6 of the *UNESCO Universal Declaration on Cultural Diversity* emphasizes on the equitable access to culturally diverse multilingual contents with help of digital technologies. Modern information and communication technologies (ICT), including internet technologies, have tremendous potentials to act as enabler for intercultural dialogue through digital dissemination of cultural information, particularly with culturally diverse contents. Cultural informatics can also bridge linguistically diverse contents through translations and adaptations. Thus, cultural informatics can help in making culture a factor of development.

UNESCO and its member states adopted an action plan for the implementation of the *UNESCO Universal Declaration on Cultural Diversity*. Some of main lines of the action plan embraced digital technologies for strengthening the access to diverse cultural resources available across the country. Text Box 1.2 provides a list of selected main lines of action plan related to access and dissemination, for implementation of the Declaration. As a member state of UNESCO, India is committed herself to take active role in the main lines of action plan for implementation of the Declaration [MCIT, 2004]. Several digitization and digital library projects in India, as documented in a recent UNESCO publication “Open Access to Knowledge and Information: Scholarly Literature and Digital Library Initiatives – the South Asian Scenario” and indicated in Table 1, have created an appropriate atmosphere for intercultural dialogue and intercultural partnership [Das, 2008]. These public-funded initiatives have tried to digitally include different communities in India, thus, further enhancing scope of cultural diversity and cultural pluralism in India [Ghosh, 2007].

The digital library initiatives in India have tried to contribute towards achieving multicultural and cross-cultural dialogs in a democratic society, in addition to making the endangered documentary resources digitally available. Digitization of documentary heritage collections in a culturally rich and diverse country is a major challenge to the ICT professionals and policymakers, due to nature of vastness versus available financial resources and institutional frameworks [UNESCO, 2009]. Thus, scaling up is real concern in India that needs to involve all possible stakeholders as well as end users.

#### Text Box 1.1: Article 6 of the UNESCO Universal Declaration on Cultural Diversity

##### ***Towards Access for All to Cultural Diversity***

“While ensuring the free flow of ideas by word and image care should be exercised that all cultures can express themselves and make themselves known. Freedom of expression, media pluralism, multilingualism, equal access to

art and to scientific and technological knowledge, including in digital form, and the possibility for all cultures to have access to the means of expression and dissemination are the guarantees of cultural diversity.”

Text Box 1.2: Select Main Lines of Action Plan for the Implementation of the *UNESCO Universal Declaration on Cultural Diversity*

- Encouraging "digital literacy" and ensuring greater mastery of the new information and communication technologies, which should be seen both as educational discipline and as pedagogical tools capable of enhancing the effectiveness of educational services.
- Promoting linguistic diversity in cyberspace and encouraging universal access through the global network to all information in the public domain.
- Countering the digital divide, in close cooperation in relevant United Nations system organizations, by fostering access by the developing countries to the new technologies, by helping them to master information technologies and by facilitating the digital dissemination of endogenous cultural products and access by those countries to the educational, cultural and scientific digital resources available worldwide.
- Ensuring protection of copyright and related rights in the interest of the development of contemporary creativity and fair remuneration for creative work, while at the same time upholding a public right of access to culture, in accordance with Article 27 of the Universal Declaration of Human Rights.
- Ensuring protection of copyright and related rights in the interest of the development of contemporary creativity and fair remuneration for creative work, while at the same time upholding a public right of access to culture, in accordance with Article 27 of the Universal Declaration of Human Rights.

## 2. Digital Library Development in India

Several scholars evaluated growth and evolution of digital library initiatives in South Asia [Das, 2008; Ghosh, 2007; Mittal, 2008; Mahesh, 2008]. Many of them feel that availability of scholarly e-resources supplements the universal access to knowledge. While commercial publishers had started offering e-contents to their subscribers, leading institutions in the world then started archiving important documents in online repositories and digital archives. Vast pool of knowledge available in the physical forms of manuscripts, rare books, out-of-print books and archival materials can be converted into digital objects and shared to a networked community using an online platform.

Development of Indian digital libraries and online repositories were planned when internet got penetrated into renowned academic and research institutions in 1990s. Indian institutions, such as Indian Institute of Science Bangalore also planned their open access digital repositories in late 1990s [Ghosh, 2007]. At the same time, Universal Digital Library was planned at global level to make available a million books online in free access mode. Its national surrogate Digital Library of India was also planned during the same time. Other digital libraries were also opened up to expand access to culturally-rich and linguistically-diverse contents to promote inter-cultural dialogues and to preserve literary assets of local and learned communities. In the last decade, many state and non-state agencies have taken significant initiatives in digitization and preservation of documentary heritage collections available across the country and also with neighbouring South Asian countries.

Table 1 shows an indicative list of multilingual digital library initiatives in India. Some of them are early starters in offering online access to Indian documentary heritage collections to global communities. Protecting and safeguarding documentary heritage of local communities and long-term preservation of endangered documentary collections are main objectives of these digital library projects.

These digital libraries could also sensitize policymakers in taking affirmative actions in content localization. Thus, many Indic language literatures got digitized and archived in searchable of digital libraries in India. Indian Ministry of Communications and Information Technology (MCIT) has established a National *Digital Libraries* Cell within the Department of Electronics and Information Technology (DeitY) to streamline development of digital libraries in India. Indian Ministry of Culture has also been involved in supporting digitization of cultural contents including manuscripts and archival materials. Majority of digital library initiatives in India, as indicated in Table 1, are public funded and targeted at common citizens.

Table 1 highlights two multimedia digital library initiatives of the Indira Gandhi National Centre for the Arts (IGNCA), funded by MCIT, namely Kalasampada: Digital Library Resources for Indian Cultural Heritage (DL-RICH) and National Databank on Indian Art and Culture (NDBIAC).

Archives of Indian Labour (AIL) is a multimedia digital library, initiated by V. V. Giri National Labour Institute and Association of Indian Labour Historians. It has very unique collection of documents of Indian labour movements that include reports, photographs, and oral testimonials.

Panjab Digital Library (PDL) is a global initiative of Nanakshahi Trust and other institutions to archive culturally rich collections of Punjabi literature and historic documents on Sikhism.

Digital Repository of West Bengal Public Library Network (DR-WBPLN) is an important multilingual digital library, established using popular content management software (CMS) DSpace. It provides access to more than 10,000 digitized rare books and government reports. Majority of them are written in Bengali language.

Muktabodha Digital Library is an initiative of Muktabodha Indological Research Institute to preserve rare Sanskrit manuscripts and texts in multiple digital formats, and make them worldwide accessible through its website for study purpose. Currently it has texts from the Kashmir Shaivism, Tantra Shastra, Kaula-Trika, Saiva-Siddhanta, Virashaiva, Pancaratra, Shree Vidya, Shakta and Natha Yoga schools.

Traditional Knowledge Digital Library (TKDL) is a flagship initiative of Council of Scientific and Industrial Research (CSIR). It attempts to curb bio-piracy and misappropriation of formulations of Indian systems of medicine (Ayurveda, Yoga, Naturopathy, Unani and Siddha). Access to TKDL is given to patent offices around the world, for sensitizing them on the prior art in formulations of Indian systems of medicine.

Table 1 also shows names of some educational digital libraries, such as National Science Digital Library (NSDL) and Vigyan Prasar Digital Library (VPDL). NSDL is an initiative of NISCAIR, New Delhi (National Institute of Science Communication and Information Resources) and VPDL is an initiative of Vigyan Prasar. Both institutions publish books on popular science and science education. NSDL and VPDL provide access to digitized e-books of popular science, published by NISCAIR and Vigyan Prasar respectively. NSDL additionally generated contents on science education topics. VPDL provides access to digitized e-books in English as well as in Hindi language.

For the convenience of this study, five important digital library initiatives are briefly evaluated in the following sections. Selected initiatives are DLI, DL-RICH, PDL, AIL and DR-WBPLN. They represent linguistically and culturally diverse documentary heritage collections. All of them maintain

open access web portals to facilitate web-based visitors make use of these unique collections. This study also identifies technical challenges common users face during their quest for knowledge discovery using these platforms.

Table 1: Indicative List of Multilingual Digital Library Initiatives in India

Name of the Initiative	Implementing Agency	Funding Agency	Website
Digital Library of India (DLI)	Indian Institute of Science; IIT Hyderabad; C-DAC	MCIT and others	<a href="http://www.new1.dli.ernet.in">http://www.new1.dli.ernet.in</a> ; <a href="http://www.new.dli.ernet.in">http://www.new.dli.ernet.in</a> <a href="http://dli.cdacnoida.in">http://dli.cdacnoida.in</a> ;
Kalasampada: Digital Library Resources for Indian Cultural Heritage (DL-RICH)	IGNCA	MCIT	<a href="http://www.ignca.nic.in/dlrich.html">http://www.ignca.nic.in/dlrich.html</a>
National Databank on Indian Art and Culture (NDBIAC)	IGNCA	MCIT	<a href="http://ignca.nic.in/ndb_0001.htm">http://ignca.nic.in/ndb_0001.htm</a>
Kritisampada : National Database of Manuscripts	National Mission for Manuscripts, IGNCA	Ministry of Culture	<a href="http://www.namami.org/pdatabase.aspx">http://www.namami.org/pdatabase.aspx</a>
Panjab Digital Library (PDL)	Panjab Digital Library	Nanakshahi Trust and others	<a href="http://www.panjabdigilib.org/">http://www.panjabdigilib.org/</a>
Digital Repository of WBPLN (DR-WBLLN)	West Bengal Public Library Network (WBPLN), CDAC Kolkata	Directorate of Library Services, West Bengal	<a href="http://dspace.wbpubliclibnet.gov.in/dspace/">http://dspace.wbpubliclibnet.gov.in/dspace/</a>
Archives of Indian Labour (AIL)	V. V. Giri National Labour Institute & Association of Indian Labour Historians	Ministry of Labour	<a href="http://www.indialabourarchives.org/">http://www.indialabourarchives.org/</a>
Muktabodha Digital Library	Muktabodha Indological Research Institute	Donations from Individuals & Trusts	<a href="http://muktalib5.org/digital_library.htm">http://muktalib5.org/digital_library.htm</a>
Traditional Knowledge Digital Library (TKDL)	Council of Scientific and Industrial Research (CSIR)	Department of Ayurveda, Yoga & Naturopathy, Unani, Siddha and Homoeopathy (AYUSH)	<a href="http://www.tkdlib.res.in/">http://www.tkdlib.res.in/</a>
National Science Digital Library	NISCAIR, India	Council of Scientific and Industrial Research (CSIR)	<a href="http://nsdl.niscair.res.in/">http://nsdl.niscair.res.in/</a>
Vigyan Prasar Digital Library	Vigyan Prasar, India	Department of Science and Technology	<a href="http://www.vigyanprasar.gov.in/digilib/">http://www.vigyanprasar.gov.in/digilib/</a>

## 2.1 Digital Library of India Project

Digital libraries development in Indian sub-continent got accelerated with the initiation of Universal Digital Library (UDL) or Million Books Project (<http://www.ulib.org/>). The Universal Digital Library is a global collaborative project, initiated in 2001 by the Carnegie Mellon University and its partner institutions in China and India. Initially, UDL got support from the US National Science Foundation. Later, each participating country secured funding from their respective national government. In India, Indian Institute of Science, Bangalore in collaboration with Centre for Development of Advanced Computing (C-DAC), and International Institute of Information Technology (IIIT), Hyderabad initiated Digital Library of India (DLI) project in 2002 as spin-off of Universal Digital Library project. The Ministry of Communications and Information Technology (MCIT) of Government of India extended full financial support for this initiative. DLI project started establishing Regional Mega Scanning Centres (RMSCs) and other scanning centres across the country for scaling up digitization of rare books, rare periodicals and other literature, including manuscripts and copyright-free or out-of-print books. DLI established five RMSCs across the country at Hyderabad (IIIT Hyderabad), Kolkata (C-DAC Kolkata), Allahabad (IIIT Allahabad), NOIDA (C-DAC Noida) and Bangalore (IISc Bangalore).

All RMSCs started networking with source libraries in their localities for obtaining the books required for digitization. Majority of books selected for digitization were Indic language books published in all official languages of India, including in English. As on 31<sup>st</sup> August 2012, contents available in top six languages are respectively English, Sanskrit, Hindi, Telugu, Bengali and Urdu covering about 91.3% of books available with DLI web portal at [www.new1.dli.ernet.in](http://www.new1.dli.ernet.in).

DLI maintains two web portals, one at IISc Bangalore ([www.new1.dli.ernet.in](http://www.new1.dli.ernet.in) and its mirror [www.new.dli.ernet.in](http://www.new.dli.ernet.in)) and one at C-DAC Noida (<http://dli.cdacnoida.in>). Earlier, IIIT Hyderabad maintained another DLI web portal (<http://dli.iiit.ac.in>). Later this portal merged with DLI [www.new1.dli.ernet.in](http://www.new1.dli.ernet.in) site.

Scholars associated with DLI project, indicated that DLI developed a significant amount of digitized Indic language contents, covering all major Indian language. Thus, DLI becomes a testbed for Indian language technology, facilitating development of OCR (optical character recognition), TTS (text-to-speech) and other related software for Indian language computing [Balakrishnan, 2006].

In a digitization project, the scanned pages are stored in image format. Images of textual pages can be converted into computer readable, full-text searchable and editable textual documents, if any OCR software is readily available for the respective language. Rate of accuracy in commercially available OCR software is usually above 90% for standard printed texts [Wikipedia, 2012]. In DLI portals, Indian language full-text contents are available in TIFF image format only, not in textual format. This is because no OCR software was used for converting Indic language texts into textual format, due to non-availability of efficient OCR software for Indian contents. But, documents in English and other European languages were converted into textual format. Pages of English language books are stored and retrieved in image as well as textual formats. However, research team of DLI project has tried development of OCR technologies for Indic languages on experimental basis using corpora of scanned images of Indian language documents. C-DAC, one of DLI project partners, has recently launched Chitrakan OCR software for Devanagari and Bengali scripts, covering Hindi, Sanskrit, Marathi, Nepali, Bengali, Manipuri and Assamese languages. Chitrakan is the first OCR system for Indian Languages. This software is now expanding to other Indian languages as well.

### 2.1.1 Retrieving Documents from DLI portals

DLI portals provide browsing as well as searching facility from its homepage. Online database search can be performed using author, range of years, subject, language, and name of scanning centre. A user can use a single search term or a combination of search terms, such as author and language. DLI does not provide metadata in an Indian language. Metadata for Indian language book is entered in transliterated form. For, example इतिहास is rendered as Itihas in the database. User has to enter words in transliterated form in the search box to perform the search. Then, a search result is displayed with titles of books and other information. When user selects a title of the book, metadata of that book is displayed onscreen. Then he can read the book online, page by page, as tiff, or rtf, or txt or html file. As Indian language books are not OCRed, only tiff image can be viewed onscreen. For English language books, tiff, or rtf, or txt or html file for each page will be displayed and can be browsed page by page.

Similarly, a user can browse books from homepage of DLI portal by selecting title beginning with a letter (A to Z), or author's last name (A to Z), or range of years, or a subject, or a language or a source library. Finally, user will reach the book of his choice and metadata of that book is displayed onscreen. Then he can read the book online, page by page, as tiff, or rtf, or txt or html file.

DLI portal requires a plugin to view a document page-by-page that makes reading books online a difficult proposition to occasional users of DLI portal. Modern, tech-savvy young users of DLI sometimes feel frustrated and critical to its presentation of page-wise view. It is contrary to present standard of reading e-books or chapters of an e-book online in PDF format without time-consuming process of downloading each and every page of a book. Users also feel temptation to download books and read at leisure, even using a tablet computer or a laptop. A user has recently developed a downloader software named “@ABS DLI Downloader” (present version 2.2) for downloading all tiff pages of the book from DLI portal in a batch to user's computer for reading in a sequential manner [Shukla, 2009]. This saves the time of end user and helps him read the book at his convenience, without compromising with time for re-downloading.

A user feels that DLI provides shabby interface for reading books online. Its required plugins, namely Alternatiff/ Plugger plugin (for BookReader-1) and QuickTime plugin (for **BookReader-2**), **if not available with them**, also impede many users to read books at leisure [Nadig, 2007].

## 2.2 Kalasampada: Digital Library Resources for Indian Cultural Heritage (DL-RICH)

DL-RICH is an attempt to digitize contents of cultural resources available with IGNCA and its partner institutions across the country. IGNCA's mandate is to prepare national inventory and documentation of different areas of cultural studies, such as, intangible cultural heritage, tangible cultural heritage, performing arts, folk arts, manuscript traditions, handicrafts and Indology. It has three regional centres across India at Bangaluru for southern region, Guwahati for north-eastern region and Varanasi for northern region. IGNCA and its regional centres partnered with local scholars engage in documenting endangered cultural resources to unearth cultural resources hidden with local communities. Thus, IGNCA has developed a vast collection of cultural resources, including rare books, manuscripts, old photographs and handicrafts.

DL-RICH was a first phase of digitization project of IGNCA to cover its existing collections as well as collections of its partner institutions. Over time, DL-RICH became a multimedia and multilingual digital library covering texts, images, audio and video recordings, 3D artefacts. DL-RICH received

funding support from the MCIT. Majority of DL-RICH collection is accessible in offline mode within the premises of IGNCA and its regional centres. Only a part of its digitized collection is made freely accessible through DL-RICH web portal. This portal provides access to different segments of its collection with English interface and English transliterated metadata information. Further, a part the DL-RICH collection also got translated into Hindi and metadata information got transliterated into Hindi to provide access to selected collection online through the interface of CoIL-Net (Content Development in Indian Language Network), a project sponsored by MCIT. This CoIL-Net collection recently went offline temporarily, probably for maintenance work.

### 2.2.1 National Databank on Indian Art and Culture (NDBIAC)

IGNCA started second phase of digitization project as National Databank on Indian Art and Culture (NDBIAC), supported by MCIT and Archeological Survey of India (ASI). NDBIAC provides access to digitized images and audio-visuals provided by ASI and state archaeology departments. A user can access information on archaeological sites in different states of India. Archaeological site information usually contains a brief description, images and video clips. NDBIAC will provide information on all major archaeological sites in India. NDBIAC also gives access to virtual walkthroughs of archaeological monuments, presently seven. The number will be increased in coming years.

In this phase of project IGNCA has started digitizing ASI publications and rare books from ASI Library. Some of the important digitized works include all back issues of ASI journal “Indian Archaeology - A Review”, ASI reports, rare books in Indic languages (Hindi and Sanskrit) and English. NDBIAC provides free access to many digitized books and reports from its web portal <http://www.ignca.nic.in/asp/searchBooks.asp>.

### 2.2.2 National Database of Manuscripts ‘Kritisampada’

IGNCA is also hosting National Mission for Manuscripts (NMM), a national apex body for preservation and conservation of manuscript resources in the country. NMM maintains a National Database of Manuscripts named ‘Kritisampada’, an outcome of its nation-wide inventory survey of manuscripts. NMM is also working towards development of a national digital library of manuscripts. NMM has identified 45 collections of **Manuscript Treasures of India (MTI)**, or *Vijñānanidhi*. *These are very unique and rare collections of manuscripts. MTI will be given first preference in the next phase of digitization and will be digitally archived through NMM’s national digital library of manuscripts. List of 45 MTI collections is available at* [www.namami.org/manuscript%20Treasures.htm](http://www.namami.org/manuscript%20Treasures.htm).

NMM also maintains a network of manuscript repositories for widening bibliographic control and safeguarding manuscript collections available with local communities and institutions across the country. NMM partners are classified as Manuscript Resource Centres (MRCs), Manuscript Conservation Centres (MCCs), Manuscript Partner Centres (MPCs) and Manuscript Conservation Partner Centres (MCPCs). NMM established a network of 47 MRCs, 32 MCCs, 32 MPCs and more than 200 MCPCs across the country for identifying, inventorying, preservation and conservation of endangered documentary heritage collections available in the form of manuscripts.



Table 2: Inscription of India's Documentary Heritage Items on MoW Register

Name of Item	Year of inscription	Host Institution	Role with NMM	Whether Further Action Taken
<i>The I.A.S. Tamil Medical Manuscript Collection</i>	1997	<i>Institute of Asian Studies, Chennai</i>	<i>MTI, MPC</i>	<i>Microfilmed &amp; Digitized; Published books</i>
Archives of the Dutch East India Company	2003	National Archive of the Netherlands	Nil	Microfilmed & Digitized; Published books
Saiva Manuscript in Pondicherry	2005	French Research Institutions in Pondicherry	MTI, MRC	Microfilmed & Digitized; Parampara CD-ROM, Digital Library; paper transcripts
Rigveda	2007	Bhandarkar Oriental Research Institute, Pune	MTI, MRC	Microfilmed & Digitized for onsite consultation
Tarikh-E-Khandan-E-Timuriyah	2011	Khuda Bakhsh Oriental Public Library, Patna	MTI, MRC, MCC	Microfilmed & Digitized for onsite consultation
Laghukālacakratant rarājatikā (Vimalaprabhā) [Buddhist Tantric literature]	2011	Asiatic Society, Kolkata	MTI, MPCC	Microfilmed & Digitized for onsite consultation

Indian government has played active role in bringing in international recognition of Indian documentary heritage collections and more specifically endangered manuscripts collections. Almost every year, India government nominates endangered and important documentary heritage collections for inscription on the Memory of the World Registers. A list of India's nominated collections that got selected for inscription is indicated in the Table 2. After inscription, the nominating institutions have started digital preservation of their commendable documentary collections. Most of these digitally preserved collections are kept in computer databases in local servers within the institutions, or stored on CD-ROMs, and DVDs. These are kept for onsite consultation by the privileged scholars who can afford to visit those institutions. Only two of them are partially available in a public domain digital library, namely, Saiva Manuscript in Pondicherry and I.A.S. Tamil Medical Manuscript Collection, which are partially available on the Muktabodha Digital Library ([http://muktalib5.org/digital\\_library.htm](http://muktalib5.org/digital_library.htm)). All institutions have preferred the medium of microfilming for long-term preservation of these collections. However, their availability remains limited to institutions' time and space. Publishing them as printed books is also an option that was exercised by a few institutions.

Interestingly, NMM has partnership with all Indian institutions mentioned in Table 2, with playing role of a Manuscript Resource Centre (MRC), Manuscript Conservation Centre (MCC), Manuscript Partner Centre (MPC), or Manuscript Conservation Partner Centre (MCPC). All inscribed items of MoW Register from India, except the Archives of the Dutch East India Company, have been recognized by NMM as **Manuscript Treasures of India (MTI)**.

### **2.3 Multilingual Digital Libraries having Indic Language Digital Collections**

Many institutions have attempted designing and development of digital archives profiling their unique collections of documentary resources. These institutions possess a significant number of rare and out-of-print books, periodicals and old photographic collection. They identify valuable source materials from their in-house collections for digitization. Sometimes they network with other libraries and document repositories to select source materials for digitization.

Table 3 shows an indicative list of digital libraries and digital repositories having collection of Indic language documents. This Table also indicates that some digital libraries use content management software (CMS), such as DSpace and Greenstone. DSpace and Greenstone are popular free and open source software (FOSS) for building digital libraries. CMS helps in developing a structured digital library with standard sets of metadata and facility of cross-searching by external metadata harvesters. Many Indian institutions have developed institutional repositories using DSpace and EPrints software.

Digital Repository of West Bengal Public Library Network (DR-WBPLN) is a joint effort of public libraries of West Bengal under the Directorate of Library Services and C-DAC Kolkata. This repository was created at the 'Heritage Preservation Unit' of West Bengal State Central Library in technical collaboration with C-DAC Kolkata. C-DAC has digitized significant number of rare books, available with State Central Library and its associated libraries. This portal is developed using DSpace software. Full-text contents of this repository are available in PDF format. This repository also provides access to digitized government publications and gazettes. A partial set of metadata information is available both in Bengali language and English, e.g., Title of book. Availability of metadata information set in Indic language is indicated in Table 3. Other metadata information is available in English only, e.g., name of author, name of publisher. This repository helps in outreaching Bengali literature to global communities of scholars and enthusiastic book readers.

Punjab Digital Library (PDL) is a significant digital library initiative jointly created by Punjabi communities of India and abroad. The mission of PDL is to locate, digitize, preserve, collect and make accessible the accumulated wisdom of the Panjab region. PDL has been digitizing significant number of rare books, magazines, newspapers, photographs and manuscripts available with Punjabi communities in the state of Punjab and Sikh religious institutions (Gurdwaras). PDL is freely accessible to the registered users from its web portal at [www.panjabdigilib.org](http://www.panjabdigilib.org). The documents can be browsed or searched after selecting its category, namely, manuscripts, books, magazines, newspapers and photographs. Full-text contents of this digital library are available in image format and can be viewed page-wise. Table 3 also indicates that metadata information of PDL documents is available in English in transliterated form. This digital library helps in outreaching Gurmukhi and Punjabi literature to global communities of scholars and vivid book readers. PDL project was bestowed the Manthan Awards in 2010 in the category of E-Culture & Heritage.

Archives of Indian Labour (AIL) is another significant initiative for building multilingual and multimedia open access repository. AIL was set up in the month of July in 1998 as a collaborative project of V.V. Giri National Labour Institute and the Association of Indian Labour Historians. This portal is

developed using Greenstone software. Presently, it provides access to 12 unique collections related to labour movements in different industries, reports of labour commissions, publications of labour unions, transcripts of recordings of oral history. AIL is freely accessible to its users from its web portal at <http://www.indialabourarchives.org/>. The documents can be browsed or searched after selecting a collection. Documents are either textual or images. Digitized full-text contents in English language of this digital repository are available in HTML format, where as digitized documents in Indic language are stored in image format and can be viewed page-wise. Table 3 also indicates that metadata information of AIL documents is available in English in transliterated form. This digital archive helps in outreaching history of South Asian labour movements and social life of industrial workers to global communities of historians and social scientists. AIL project was bestowed as the pioneering and predecessor of all recent digital library initiatives of Indian sub-continent.

Table 3 also indicates other digital repositories providing access to digitized Indian contents. Digital Repository of Vidya Prasarak Mandal provides full-text access to some Marathi books and magazines. E-Gyankosh of Indira Gandhi National Open University (IGNOU) provides full-text access to digitized course materials in Indian language, although majority of IGNOU course materials are available in English. Similarly, ASI Digital Library of Archeological Survey of India provides full-text access to selected digitized rare books of ASI Library collections. As indicated earlier ASI Digital Library is a part of NDBIAC project of IGNCA.

Table 4 indicates another set of multilingual digital libraries, established by institutions abroad in collaboration with South Asian partner institutions. These digital libraries have significant collections of digitized contents published from South Asia. These also have contents in Indic language. Majority of contents of these digital libraries is focused on South Asian studies, South Asian literature and Indology.

Table 3: Indicative List of Indian Digital Libraries having Indic Language Contents

Name of Digital Library	Organization	Indic Language Collections	CMS used	Whether Metadata in Indic Language Available	Metadata in Indic Language
Digital Repository of W.B. Public Library Network <a href="http://dspace.wbpublicnet.gov.in:8080/dspace/">http://dspace.wbpublicnet.gov.in:8080/dspace/</a>	West Bengal State Central Library & CDAC Kolkata	Digitized Rare Books (10195 items, about 80% in Bengali)	DSpace	Yes, Partial	Title, Appears in Collections, Description
Panjab Digital Library <a href="http://www.panjabdigilib.org/">http://www.panjabdigilib.org/</a>	Panjab Digital Library; Nanakshahi	Manuscripts 704; Digitized Books 994; Magazines 432; Newspapers 540; Photographs 103	-	No	-
Archives of Indian Labour (AIL) <a href="http://www.indialabourarchives.org/">http://www.indialabourarchives.org/</a>	V. V. Giri National Labour Institute & Association of Indian Labour Historians	Digitized Reports, A-V materials, Images.	Greenstone	No	-
Digital Repository of VPM <a href="http://dspace.vpmthane.org:8080/jspui/index.jsp">http://dspace.vpmthane.org:8080/jspui/index.jsp</a>	Vidya Prasarak Mandal, Thane	Marathi E-Books; Marathi E-Journals (373 items).	DSpace	Yes, Partial	Title, Authors, Publisher, Appears in Collections
E-Gyankosh <a href="http://www.egyankosh.ac.in/">http://www.egyankosh.ac.in/</a>	Indira Gandhi National Open University, New Delhi	<b>Digital learning resources</b> (25025, about 1% in Hindi and other Indic lang.).	DSpace	No	Title, Appears in Collections
ASI Digital Library <a href="http://www.ignca.nic.in/asp/searchBooks.asp">http://www.ignca.nic.in/asp/searchBooks.asp</a>	ASI Library; IGNCA New Delhi	Digitized Rare Books, ASI Publications (out-of-print), journal <i>Indian Archaeology</i> .	-	No	-
Data as on 1st September 2012					

Table 4: Overseas Digital Libraries having Sourcing Partners in South Asia/ Collections from South Asia

Name of the Initiative	Implementing Agency	Funding Agency	URL
Digital Himalaya	Digital Himalaya Project team	University of Cambridge	<a href="http://www.digitalhimalaya.com/">http://www.digitalhimalaya.com/</a>
Tibetan and Himalayan Library (THL)	University of Virginia Library; Institute for Advanced Technology in the Humanities, USA	University of Virginia	<a href="http://www.thlib.org/">http://www.thlib.org/</a>
The Digital South Asia Library	University of Chicago and the Center for Research Libraries, USA	University of Chicago	<a href="http://dsal.uchicago.edu/">http://dsal.uchicago.edu/</a>

### 3. Conclusion

Indian multilingual digital library initiatives have shown keen interests in ‘lean backward’ to digitize important documentary heritage collections available with Indian institutions. Most of these items have status of rare, out-of-print or copyright-free books and documents. After completing the process of digitization, these rare books are usually archived in an online digital library platform. Educated common people consult the open access digital libraries to read digitized books online and even download some of them for future reading. A rare book in an Indic language gets much attention to the general book readers, because readers get easy access to a full document. Popular literature if available with a digital library platform, it will increase the chance of web visibility of respective digital library. On the other hand, documentary heritage collections in the form of manuscripts are of special interests of scholars or subject specialists.

All five digital libraries evaluated in this study have sufficient propositions to become popular choices. Digital Library of India has largest digital collections and its technical glitches are also higher than other initiatives due to flaw in design. DLI also has many dead web-links showing pages do not exist while browsing, which needs to be corrected.

Present study tries to focus availability of contents of documentary heritage collections in online platforms. After evaluating national and institutional initiatives, we still believe that majority of digital library initiatives failed to archive all their digitized contents at fullest extend. While international and national policy instruments are in place to augment access to cultural diversity of a large country like India, most of the initiatives are short-focused, in terms of incorporating interoperable, cross-searching and metadata harvesting functionalities. Many of the online resources could not get global visibility due to low level of outreach, advocacy and awareness raising activities.

This paper indicates some digital libraries provide metadata information in Indian language or in transliterated English, for the documents in a respective Indian language. Lack of searchable metadata in Indian language hampers readers’ quest for knowledge discovery in a digital environment. It also makes documents inaccessible to some prospective readers. Gradually, we need to focus on searchable metadata information in Indian language.

Multilingual digital library initiatives in India have helped in bridging digital divide in the country by making Indian language documents freely available to the masses and pushing content localization efforts of associated online platforms.

India has gained momentum in open access movement, by establishing open access channels of digital publishing. Many authors writing in vernacular languages in India would be interested in putting documents in Indian digital libraries with creative commons licenses. Thus, we now need to focus on “lean forward” to include born digital contents in multilingual digital library collections.

## References

- Balakrishnan, N.; Reddy, Raj; Ganapathiraju, Madhavi; Ambati, Vamshi (2006) Digital Library of India: A Testbed for Indian Language Research. *TCDL Bulletin*, 3 (1).
- Das, Anup Kumar (2008). *Open Access to Knowledge and Information: Scholarly Literature and Digital Library Initiatives – the South Asian Scenario*. New Delhi: UNESCO.
- Das, Anup Kumar; Dutta, Chaitali; Sen, B.K. (2007) "Information retrieval features in Indian digital libraries: a critical appraisal", *OCLC Systems & Services*, 23(1), pp. 92-104.
- Ghosh, S.B.; Das, Anup Kumar (2007) "Open Access and Institutional Repositories – A Developing Country Perspective: a case study of India" *IFLA Journal*, 33(3), pp.229-250.
- India, Ministry of Communications and Information Technology (2004). *Digitization of Culture – Background Note for Asia IT Ministers' 2nd Summit, Hyderabad*.
- Mittal, Rekha; Mahesh, G. (2008) "Digital libraries and repositories in India: an evaluative study", *Program: Electronic Library and Information Systems*, 42(3), pp.286-302.
- Mittal, Rekha; Mahesh, G. (2008) "Digital Libraries in India: A Review", *Libri*, vol. 58, pp. 15-24.
- Nadig, Hari Prasad (2007). “Digital Library of India: Download all that you can...”. Retrieved on 1 September 2012 from <http://hpnadig.net/blog/index.php/archives/2007/02/22/download-all-that-you-can>
- Shukla, Alok Bhushan (2009). “Where Knowledge is Free – Digital Library of India”. Retrieved on 1 September 2012 from <http://alokshukla.wordpress.com/2009/12/11/where-knowledge-is-free-digital-library-of-india/>
- UNESCO (2001). *Universal Declaration on Cultural Diversity*. Retrieved 1 February 2009 from <http://www.un-documents.net/udcd.htm>
- UNESCO (2009). *UNESCO World Report: Investing in Cultural Diversity and Intercultural Dialogue*. Paris: UNESCO Publishing.
- Wikipedia (2012). *Optical character recognition*. Retrieved on 1<sup>st</sup> September 2012 from [http://en.wikipedia.org/wiki/Optical\\_character\\_recognition](http://en.wikipedia.org/wiki/Optical_character_recognition).